Detection and Transition Analysis of Engagement and Affect in a Simulation-based Combat Medic Training Environment

Jeanine A. DeFalco[1], Ryan S.J.d.Baker[1]

[1]Teachers College, Columbia University, New York, NY
jad2234@tc.columbia.edu, baker2@exchange.tc.columbia.edu

Abstract. Developing intelligent tutoring systems that respond effectively to trainee or student affect is a key part of education, particularly in domains where learning to regulate one's emotion is key. Effective affect response relies upon effective affect detection. This paper discusses an upcoming cooperative study between the Army Research Laboratory, Teachers College, Columbia University, and North Carolina State University, with the goal of developing automated detectors that can infer trainee affect as trainees learn by interacting with the vMedic system, which trains learners in combat medicine. In this project, trainee interactions with vMedic will be synchronized with observations of engagement and affect; and physical sensor data on learners, obtained through GIFT's Sensor Module. The result will be models of trainee affect, ready for integration into the GIFT platform, which can leverage sensor information when available, but which can make reasonably accurate inference even without sensor data.

Keywords: GIFT, vMedic, affect, tutoring, intelligent tutoring systems, learning, automated detectors, game-based training

1 Introduction

In recent years, there has been increasing interest in modeling affect within intelligent tutoring systems [7, 11] and using these models to drive affect-sensitive interventions [2]. In this paper, we describe an ongoing collaborative project between the Army Research Laboratory, Teachers College Columbia University, and North Carolina State University, which has the goal of developing automated detection of trainee affect that can leverage sensors when they are available, but which can function robustly even when sensors are not available.

Within this research, trainee affect will be studied in the context of the vMedic, (a.k.a. TC3Sim), a game developed for the U.S. Army by Engineering and Computer Simulations (ECS) in Orlando, Florida, to train combat medics and combat lifesavers on providing care under fire and tactical field care. Trainees will also complete material on hemorrhage control within the auspices of the GIFT framework [12], the Army Research Laboratory's modular framework for Computer-Based Training

Systems, with the goal of integrating eventual affect detection into the GIFT framework's User Module (realized as necessary within the Sensor Module). In turn, the affect detectors will be built into the pedagogies realized through the GIFT Framework's Pedagogical Module, for instance to realize interventions through the embedded instructor and other non-player characters.

In this fashion, this project will contribute not just to the assessment of affect within vMedic, but also to the GIFT framework's broader goal of integrating a range of types of models and detectors into the GIFT framework. By serving as a test case for incorporating two types of detection into GIFT (sensor-free affect detection, and sensor-based affect detection), this project will assist in understanding how GIFT needs to be enhanced to incorporate the full range of models currently being developed by this research community.

Using these detectors, further work will be conducted to study student affective trajectories within vMedic, which affective and engagement states influence learning of the key material within vMedic, and how trainee affect can best be supported based on the results of affect detection. The work to study the relationship between affect, engagement, and outcome variables will provide important evidence on which affective states and engagement variables need to be responded to in a suite of optimally effective computer-based tutoring systems for Army use. Also, integrating automated detectors and interventions into vMedic through GIFT's Trainee Module and Pedagogical Module will provide a valuable example of how to respond to trainees' negative affect and disengagement, a valuable contribution in improving vMedic and similar training systems used by the U.S. Army.

2 Previous Research: Theoretical Grounding

Affect influences learning in at least three ways: memory, attention, and strategy use [16, 18]. Overly strong affect can contribute to cognitive load in working memory, reducing the cognitive resources available to students in learning tasks [13]. Beyond this, negative affective states such as frustration and anxiety can draw cognitive resources away from the task at hand to focus on the source of the emotion [20]. These high-intensity negative affective states can be particularly important for trainees learning content that is emotionally affecting or relevant to their future goals. Combat medicine training for soldiers has each of these components; it is relevant to future situations where they or their fellow soldiers may be in physical danger, and the training in vMedic is designed to be realistic and to involve scenarios where soldiers scream in pain, for example.

However, boredom and disengagement are also relevant to trainees engaging in a task that is not immediately relevant, even if it is relevant to a trainee's longer-term goals. Boredom results in several disengaged behaviors, including off-task behavior [8] and gaming the system [5], when a student intentionally misuses the learning software's help or feedback in order to complete materials without learning. Both gaming the system and off-task behavior have been found to be associated with poorer learning in online learning environments [cf. 4].

However, automated systems that infer and respond to differences in student affect can have a positive impact on students, both in terms of improved affect and improved learning [2, 13]. Similarly, automated interventions based on engagement detection can improve both engagement and learning [2].

A key aspect of automated intervention is the need to detect differences in student affect and engagement, in order to intervene effectively. Recent work has detected these constructs, both using sensors [15], and solely from the student's interactions with the learning system [5, 7, 8]. In recent years, sensor-free models have been developed of a range of behaviors associated with engagement or disengagement: gaming the system [3, 4], off-task behavior [3], self-explanation – an engaged behavior [6], carelessness [18], and WTF ("without thinking fastidiously") behavior, actions within a learning system not targeted towards learning or successful performance [24, 34], among other constructs.

Similarly, automated detectors have been developed that can infer affect solely from student interactions with educational software [7, 10, 11]. However, better performance has typically been achieved by systems that infer affect not only from student interactions, but also from information obtained by physiological sensors. These recognition models use a broad range of physical cues ensuing from affective change. Observable physical cues include body and head posture, facial expressions, and posture, and changes in physiological signals such as heart rate, skin conductivity, temperature, and respiration [1]. In particular, galvanic skin response (GSR) has been correlated with cognitive load and stresses [15], frustration [9], and detecting multiple user emotions in an educational game [10].

3 Project Design

The first step towards developing automated detectors of student affect is to obtain "ground truth" training labels of student affect and engagement. Two approaches are typically chosen to obtain these labels: expert human coding, and self-report [11]. In this project, we rely upon expert human coding, as self-report can be intrusive to the processes we want to study, and self-presentation and demand effects are also likely to be of concern with the population being studied (military cadets are unlikely to want to report that they are frustrated or anxious).

These training labels will be collected in a study to be conducted at West Point, the United States Military Academy. Each trainee will use vMedic for one hour in a computer laboratory, in groups of ten at a time. The following sources of student data will be collected: field observations of trainee affect and engagement; the Immersive Tendencies Questionnaire (ITQ), an instrument to gauge an individual's propensity to experience presence in mediated environments a priori to system interaction; the Sense of Presence questionnaire, a 44-item questionnaire that rates subjective levels of presence on four separate factors: (1) Sense of Physical Space (19 items); (2) Engagement (13 items); (3) Ecological Validity/Naturalness (5 items); and (4) Negative Effects (6 items) [19];, a pre-and post test on hemorrhage control (a total of 12 questions, same questions used in pre-and post-test, though ordered differently), and physical sensor data for students as they play the game. The following physical sensors will be used: Q-sensors, and Kinect depth sensors. Q-sensors track skin conductance data, a measure of arousal, while Kinect depth sensors record depth-map images to support recognition of postural positions.

Within this study, expert codes of trainee affect and engagement will be collected by a trained coder (the first author) using the BROMP 1.0 field observation protocol [16]. The field observations will be conducted in a pre-chosen order to balance observation across trainees and avoid bias towards more noteworthy behaviors or affect. Observations will be conducted using quick side glances in order to make it less clear when a specific trainee is being observed. Coding includes recording the first behavior and affect displayed by the trainee within 20 seconds of the observation, choosing from a predetermined coding scheme. The affect to be coded includes: frustration, confusion, engaged concentration [5], boredom, anxiety, and surprise. Affect will be coded according to a holistic coding scheme. Behavior coding includes: on-task behavior, off-task behavior, gaming the system, "psychopath" behavior (friendly fire, killing bystanders), and WTF ("without thinking fastidiously") behavior, where the trainee's actions have no relation to the scenario [17]. In order to be BROMP-certified, a coder must achieve inter-rater reliability of 0.6 or higher to another BROMP-certified coder; two coders are currently trained at Teachers College, and are available for the project.

Field observation coding will be conducted within a handheld Android app, HART, designed for this purpose [7]. The field observations will be synchronized to the other data sources, based on use of an internet time server. Synchronization will be with reference to several data sources, including trainee interactions with vMedic, provided through the GIFT framework's Trainee Module, and physical sensor data on learners, obtained through GIFT's Sensor Module. We anticipate synchronization to involve a skew of 1-2 seconds, based on the time required to enter observations.
The GIFT platform includes a synchronization library, which connects to an Internet timeserver so that a precise time-stamp can be added to the logs of trainee interactions with vMedic, and the corresponding sensor data. By connecting to the exact same timeserver, the interactions with vMedic, field observations of engagement and affect, and physical sensor data on learners, three data sources can be precisely synchronized.

Automated detectors will be developed using the interaction logs alone, for use when physiological sensors are not available, and using the sensors as well, for situations where they are. A standard approach of conducting feature engineering and then developing classifiers, and validating the classifiers using student-level cross-validation, will be used.

4 Conclusion

The current project has the goal of enhancing the GIFT framework through the creation of models that can infer trainee engagement and affect. This project is expected to both enhance the capacities of the vMedic software, and to provide a model for how this type of detection can be integrated into the GIFT framework more generally. As such, this project is just one small component of the larger efforts that are currently being pursued by the Army Research Lab, to make the GIFT framework a general and extensible platform to achieve the US Army's overall objective of applying learning theory and state-of-the-art learning technology to achieve superior training results for warfighters [14]. We anticipate that this collaborative effort will provide useful information on the future enhancement of the GIFT platform; as such, this project represents a step towards the vision of adaptable and scalable Computer-Based Training Systems helping to enhance the training of U.S. Army military personnel and prepare U.S. soldiers for the conflicts of the future.

Acknowledgments. We thank our research colleagues at the Army Research Lab, Dr. Robert Sottilare, Benjamin Goldberg and Keith Brawner, as well as at North Carolina State University, Dr. James Lester, Dr. Bradford Mott, and Jonathan Rowe. This research is supported by a grant by the Army Research Lab. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of ARL or NCSU.

References

1. Allanson, J., Fairclough, S. H. A research agenda for physiological computing. In Interacting with Computers, 16 (2004), 857–878.
2. Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S.,Woolf. B. Preparing disengagement with non-invasive interventions. In Proceedings of the 13th International Conference on Artificial Intelligence in Education,.(2007) 195–202.
3. Baker, R.S.J.d. Is Gaming the System State-or-Trait? Educational Data Mining Through the Multi-Contextual Application of a Validated Behavioral Model. In Proceedings of the Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling 2007, 76-80.
4. Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System." In Proceedings of ACM CHI 2004: Computer-Human Interaction 2004, 383-390.
5. Baker, R.S.J.d., D'Mello, S.K., Rodrigo, .M.T., Graesser, A.C. Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. In International Journal of Human-Computer Studies, 48 (2010), 223-241.
6. Baker, R.S.J.d., Gowda, S.M., Corbett, A.T. Automatically Detecting a Student's Preparation for Future Learning: Help Use is Key. In Proceedings of the 4th International Conference on Educational Data Mining 2011, 9-188.
7. Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., Rossi, L. Sensor-free automated detection of affect in a Cognitive Tutor for Algebra. In Proceedings of the 5th International Conference on Educational Data Mining 2012, 126-133.
8. Baker, R.S.J.d., Moore, G., Wagner, A., Kalka, J., Karabinos, M., Ashe, C., Yaron, D. The Dynamics Between Student Affect and Behavior Occuring Outside of Educational Software. In Proceedings of the 4th bi-annual International Conference on Affective Computing and Intelligent Interaction 2011.
9. Burleson, W. Affective learning companions: strategies for empathic agents with real-time multimodal affect sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance. In unpublished Doctoral Dissertation. Cambridge, MA: Massachusetts Institute of Technology (2006).
10. Conati, C. Probabilistic Assessment of User's Emotions in Educational Games. In Journal of Applied Artificial Intelligence, 16 (2002), 555-575.

11. D'Mello, S.K., Craig, S.D., Witherspoon, A. W., McDaniel, B. T., Graesser, A. C. Automatic Detection of Learner's Affect from Conversational Cues. In User Modeling and User Adapted Interaction, 18 (2008), 45-80.
12. Goldberg, B., Brawner, K., Sottilare, R, Tarr, R., Billings, D., & Malone, N. Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies. In Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2012.
13. Linnenbrink, E.A., Pintrich, P.R. Multiple Pathways to Learning and Achievement: The Role of Goal Orientation in Fostering Adaptive Motivation, Affect, and Cognition. In Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance, C. Sansone & J.M. Harackiewicz, Eds. Academic Press, San Diego, 2000, 195-227.
14. McQuiggan, S., Rowe, J., Lee, S., Lester, J. Story-based Learning: The Impact of Narrative on Learning Experiences and Outcomes. In Proceedings of the Ninth International Conference on Intelligent Tutoring Systems (2008), 530-539.
15. Mohammad, Y., & Nishida, T. Using physiological signals to detect natural interactive behavior. In Spring Science & Business Media, 33(2010) 79-92.
16. Ocumpaugh, J., Baker, R.S.J.d., Rodrigo, M.M.T. Baker-Rodrigo Observation Method Protocol (BROMP) 1.0. Training Manual version 1.0. Technical Report. EdLab: New York, NY, Manila, Philippines: Ateneo Laboratory for the Learning Sciences (2012).
17. Sabourin, J., Rowe, J., Mott, B., and Lester, J. When Off-Task in On-Task: The Affective Role of Off-Task Behavior in Narrative-Centered Learning Environments. Proceedings of the 15th International Conference on Artificial Intelligence in Education, pp. 534-536, 2011.
18. San Pedro, M.O.C., Rodrigo, M.M., Baker, R.S.J.d. The Relationship between Carelessness and Affect in a Cognitive Tutor. In Proceedings of the 4th bi-annual International Conference on Affective Computing and Intelligent Interaction 2011.
19. Witmer, B.G. & Singer, M.J. Measuring presence in virtual environments: A presence questionnaire. In Presence, 7 (1998), 225-240.
20. Zeidner, M. Test Anxiety: The State of the Art. Springer, Berlin, 1998.

Authors

Jeanine A. DeFalco is a Doctoral Research Fellow in Cognitive Studies at Teachers College, Columbia University. Jeanine's research interests include embodied cognition and role-play as a methodology for improving analogic reasoning and creative problem solving in both live and simulated learning platforms. A member of Kappa Delta Pi, the international honors society for education, Jeanine has a Masters in Educational Theatre, Colleges and Communities, from the Steinhardt School, New York University, and a Masters in Drama Studies from The Johns Hopkins University. Jeanine's paper, "Cognition, O'Neill, and the Common Core Standards," has an expected publication in the *Eugene O'Neill Journal* for Fall 2013, and she will be presenting this same paper at the July 2013 American Alliance for Theatre and Education conference in Bethesda, MD. Other conference presentations include "Drama as an epistemology for pre-service teachers" forW the 2012 National Communication Association conference in Orlando, FL, and "Teaching O'Neill" at the 8th International Eugene O'Neill Conference, 2011, in New York, NY.

Ryan S.J.d. Baker is the Julius and Rosa Sachs Distinguished Lecturer at Teachers College, Columbia University. He earned his Ph.D. in Human-Computer Interaction from Carnegie Mellon University. Baker was previously Assistant Professor of Psychology and the Learning

Sciences at Worcester Polytechnic Institute, and he served as the first Technical Director of the Pittsburgh Science of Learning Center DataShop, the largest public repository for data on the interaction between learners and educational software.

He is currently serving as the founding President of the International Educational Data Mining Society, and as Associate Editor of the Journal of Educational Data Mining. His research combines educational data mining and quantitative field observation methods in order to better understand how students respond to educational software, and how these responses impact their learning. He studies these issues within intelligent tutors, simulations, multi-user virtual environments, and educational games.